

# The Real Problem with Evolutionary Debunking Arguments

## ABSTRACT

There is a substantial literature on evolutionary debunking arguments (EDAs) in metaethics. According to these arguments, evolutionary explanations of our moral beliefs pose a significant problem for moral realism, specifically by committing the realist to an unattractive pessimism about the prospects of our having moral knowledge.

In this paper I argue that EDAs exploit an equivocation between two distinct readings of their central claim. One is plausibly true but has no epistemic relevance, and the other would have epistemic consequences for realism, but is false. If I'm right, this undermines attempts to use evolutionary explanations to debunk belief in other domains too.

**Keywords:** metaethics, moral epistemology, evolutionary debunking arguments, moral realism.

## Introduction

According to a common form of argument, if there is a plausible evolutionary backstory that would explain our moral beliefs, this would make trouble for moral realists in that it would commit them to an unattractive pessimism about the prospects

of moral knowledge.<sup>1</sup> Arguments that take this form are known as evolutionary debunking arguments (EDAs).

In this paper I argue that evolutionary explanations of our moral beliefs don't in fact have any bearing on the epistemic status of these beliefs. Perhaps there are other reasons to think that realism has trouble accounting for moral knowledge, but evolutionary considerations - I argue - are not among them. EDAs, I argue, exploit an equivocation between two distinct readings of their central evolutionary claim. One is plausibly true but has no epistemic relevance, and the other would have epistemic consequences for realism, but is false.

I begin by saying in more detail what realism is, what EDAs are, and identifying the ambiguity I take them to exploit. In §2-§3 I consider in detail a number of different kinds of EDA, and argue that each one rests on the same equivocation. In §4 I address objections. I also address a recent paper in the literature on EDAs that comes close to recognising the problem I set out here, and outline two important respects in which the problem I identify here is a good deal more severe.<sup>2</sup>

---

<sup>1</sup> See Street 2006; Kahane 2011; Ruse & Wilson 1986; Joyce 2006a, 2006b; Greene 2008; Lillehammer 2003; Levy 2006. For responses see Clarke-Doane 2012; Enoch 2010; FitzPatrick 2014; Graber 2012; Shafer-Landau 2012; Skarsaune 2011; Wielenberg 2010.

<sup>2</sup> Mogensen 2015. The arguments in the present paper were developed in ignorance of Mogensen's paper.

# Section 1

## 1.1 Realism

EDAs take as their target the metaethical position known as moral realism. By ‘moral realism’, most participants in the debate have in mind a view that comprises at least the following tenets:

R1. Moral statements have truth-conditions

R2. At least some moral statements are true

R3. These truth-values are (relevantly)<sup>3</sup> independent of the attitudes of (even idealised)<sup>4</sup> agents.<sup>5</sup>

Some philosophers discussing EDAs have concentrated on more demanding conceptions of realism, possibly because they think that the less demanding versions will have easier ways out of the problems posed by EDAs. David Enoch, for example, takes the non-naturalist brand of realism that he favours to be the version of realism against which the debunkers have the best chance of success.<sup>6</sup> The debunkers’ target would then include a fourth tenet, that moral claims are not reducible to natural ones. My thesis here will be that evolutionary considerations don’t pose problems for *any* of these theories in the way that the debunkers have claimed. If what I go on to argue in

---

<sup>3</sup> Most realists who include a mind-independence clause don’t want to deny *all* kinds of dependence of normative facts on human attitudes. They will want to allow that the wrongness of a sharp remark or an ill-judged joke can depend on the fact that it caused someone to feel hurt or embarrassed. For discussion of this see Shafer Landau 2003: 15. See also Clarke-Doane 2012: 316-7.

<sup>4</sup> I’m following Shafer Landau 2012 in adding this qualification.

<sup>5</sup> Street 2006: 110-111.

<sup>6</sup> Enoch, *Taking Morality Seriously*, pp. 151-184. See also Street 2006: 111-112; Graber 2012: 590.

this paper is correct, then, even the most demanding variants of realism face no particular problem from evolutionary considerations. Readers who are of the view that some versions of realism escape EDAs while others do not, should take my use of ‘realism’ here to denote the version of realism that they take to be the easiest target for the debunkers, and should take my claim to be that even this survives the debunkers’ attacks.

## 1.2 EDAs

EDAs all begin with the claim that our moral beliefs have been in some sense shaped by natural selection. Call this the Evolutionary Backstory (EB) Claim. Most debunkers take this shaping to happen indirectly: they aren’t claiming that the beliefs themselves are inherited from our ancestors; rather, they hold that these beliefs are the product of heritable dispositions.<sup>7</sup> As one influential debunker puts it, dispositions to experience certain kinds of behaviour as *called for*, such as looking after one’s children, or reciprocating kindness, underlie our moral beliefs.<sup>8</sup> These dispositions, the debunkers claim, are the product of natural selection: the reason these dispositions, and not others, got passed down is that these dispositions are fitness-enhancing. Creatures who have these dispositions are, *ceteris paribus*, more likely to survive and to reproduce than creatures who lack them. Since natural selection explains why we have these dispositions, and these dispositions underlie our moral beliefs, natural selection indirectly explains our moral beliefs. We have the moral beliefs we have because these are the ones that go with the fitness-enhancing dispositions.

---

<sup>7</sup> See, for discussion of this, Street 2006: 118-120.

<sup>8</sup> See Street 2006: 119; Shafer-Landau, 2012: 1.

But if realism is true, there is good reason to think that the fitness-enhancingness of these beliefs did not depend in any way on their being true. For realists, true moral beliefs are those that faithfully represent mind-independent moral reality. But even if there are mind-independent moral facts, it looks implausible that the fitness-enhancingness of our moral beliefs would depend on how well they latch onto those facts. The realist should admit, it seems, that the fitness-enhancingness of a moral belief is an entirely separate issue from its accuracy. But add this to the EB claim, and it appears we have a backstory for how we came to have our moral beliefs in which - the realist has to admit - truth plays no role.

The reasoning, then, is this:

- (1) **EB Claim:** Our moral beliefs have been shaped by natural selection
- (2) Natural selection is indifferent to the truth of these beliefs
- (3) Our moral beliefs have been shaped by forces that are indifferent to their truth (from 1, 2)<sup>9</sup>

Debunkers take (3) to be epistemically worrying for the moral realist in two broad kinds of way: if truth plays no role in our acquisition of our moral beliefs, then:

- (i) It's very unlikely that these beliefs are going to be true ones: for them to be true, it would take an enormous coincidence, and

---

<sup>9</sup> See Kahane 2011 esp. p. 114, Street 2006, esp. pp. 125-134; Street 2008 esp. pp. 208-9; Joyce 2006a: 222; Joyce 2006b, esp. pp. 135-6. See also Clarke-Doane 2012: 315-7; Shafer-Landau 2012, esp. p. 3.

- (ii) Even if, by some massive fluke, these beliefs were true, the flukiness of our getting it right would surely stop these beliefs counting as knowledge (or indeed, having any other kinds of epistemic virtues: justification, reliability, etc.).

Moral realism, either way, is committed to an unattractive pessimism about the prospects of moral knowledge.

Although debunkers don't tend to distinguish between (i) and (ii), they are in fact two distinct (albeit related) kinds of worry for the moral realist. Call arguments of the first kind *Truth-Undermining Arguments*, and arguments of the second kind *Warrant-Undermining Arguments*. I argue in what follows that both kinds of argument are guilty of the same mistake. Warrant-undermining arguments (and less obviously, I will argue, truth-undermining arguments) rely on taking (1)-(3) to show that there is something epistemically wrong with our moral beliefs. But, I will argue, (1)-(3) in fact fail to show this. The root of the problem is premise 1, the EB claim. Perhaps 2 is also questionable, but I accept 2 for argument's sake.<sup>10</sup> The EB claim, I argue, is ambiguous between two different readings, one of which is plausible but epistemically irrelevant, the other of which would potentially be epistemically relevant, but is implausible:

**The *predicative reading*** of the EB claim takes it to be predicating something of each individual. On this reading, it is saying that you, and I, and everybody else, have each individually been caused to accept the moral beliefs we accept by the forces of natural selection.

---

<sup>10</sup> For critiques of EDAs that challenge premise 2, see Enoch 2010, Wielenberg 2010.

**The *quantificational reading*,** on the other hand, takes the EB claim to be a very different kind of claim, that the fact that the world contains many individuals who have these moral beliefs, and not many (if any) who don't, is explained by natural selection (because individuals who had those kinds of beliefs generally had more offspring, leading to a situation where more and more individuals with these beliefs exist, and individuals without them make up a smaller and smaller proportion of the population, eventually leading to a situation where all (or the vast majority) of the individuals who exist have these beliefs).

I call this the quantificational reading because what is being explained is a quantified claim. Quantifiers include not just 'all' and 'some', but also 'many' and 'most', so even if you think that what's being explained is not that *all* humans have these beliefs, but only that *many* or *most* humans do, it's still a quantified claim that is being explained on this reading of the EB claim.

Ambiguities between predicative and quantificational readings are not unique to the EB claim. In fact, it's plausible that any time we talk about explaining why members of a group have certain traits, we get this kind of ambiguity. Consider this case from Elliot Sober. Suppose all the children in a given classroom are able to read at Grade Three level. Now consider the question of what explains this. This question is ambiguous in the following way. In one sense, pointing out that the ability to read at Grade Three level is a criterion for being admitted to that classroom would be an explanation of the fact that all the children in that classroom can read at Grade Three level; any child who is unable to read at Grade Three level would not be admitted, and this is why all the children there can read at that level. In another sense, explanation of the fact that all the children in the classroom are able to read at Grade

Three level would involve showing why each of the individuals who are in fact in the classroom are able to read at Grade Three level. Suppose the children in the classroom are Katie, Alexander, Sally, and James. If we can explain why Katie can read at Grade Three level, why Alexander can, and so on, then we have (in a different sense) an explanation of why the children in that classroom can read at Grade Three level.<sup>11</sup>

Note that first kind of explanation doesn't leave us any the wiser about how Katie, Alexander, and the others got to Grade Three level for their reading. It offers no causal story about how they came to have that ability. It is an explanation in the quantificational sense, of explaining why the classroom contains only children who have a given trait, not in the predicative sense of explaining why these particular children have this trait.

Here's another example: Sarah is an interior designer. All her clients are based in New York's upper East side. In one sense, an explanation for why all her clients are based there is that she only advertises there. This is an explanation in the quantificational sense: it explains how her clientele came to be composed only of people who live in the upper East side. But it is not an explanation in the predicative sense: it doesn't say anything about why each of these individuals who are in fact her clients are based there.

Which reading of (1) we accept determines where (1)-(3) gets us. In particular, different readings of (3) are licensed depending on how (1) is taken. If (1) is read in the

---

<sup>11</sup> Sober 1984: 149.



predicative way, a predicative reading of (3) would follow from (1) and (2), giving us the following argument:

#### Predicative Reading of (1)-(3)

- (1) **Predicative EB Claim:** you, and I, and everybody else, have each individually been caused to accept the moral beliefs we accept by the forces of natural selection.
- (2) Natural selection is indifferent to the truth of these beliefs
- (3) Forces that are indifferent to moral truth have made it the case that I have these moral beliefs, and you do, and that other individuals do etc.

If (1) is read in the quantificational way, however, (3) would also need to be read quantificationally for the reasoning to be valid. The quantificational reading of the argument would be as follows:

#### Quantificational Reading of (1)-(3)

- (1) **Quantificational EB Claim:** Natural selection has made it the case that individuals with these moral beliefs exist, rather than individuals with different moral beliefs.
- (2) Natural selection is indifferent to the truth of these beliefs
- (3) Forces that are indifferent to moral truth have made it the case that individuals with these moral beliefs exist, rather than individuals with different moral beliefs.

You can see how the predicative reading of (3) would be potentially epistemically worrying for the moral realist. In general if an individual's belief that P is a result of forces that are indifferent to whether P is true, then this is normally taken to be a good case for thinking that this belief is not knowledge. The quantificational reading, on the other hand, does not obviously lead to epistemic worries. While the backstory of how an *individual* came to believe that P is potentially relevant to the epistemic status of her belief, it's not clear that the backstory of how *the world came to contain* individuals who believe that P would be relevant.

Unfortunately for the debunker, however, it's hard to see why we should take the predicative reading to be plausible. First, the debunker hasn't given us any reason to think the predicative reading is true. What debunkers tend to say in support of the EB claim is support for the quantificational reading. For example, Sharon Street says: 'It is fairly obvious why, other things being equal, ancestors with these evaluative tendencies would have left more descendants than counterparts who, for example, viewed their survival as bad, their children's lives as worthless, or the fact that someone has helped them as a reason to hurt that person in return'.<sup>12</sup>

Second, claims about evolutionary backstories are, in general, quantificational claims. When we talk about a particular trait common to members of a given species being the result of natural selection, we are saying, roughly, that individuals who had that trait or tendencies towards it had more offspring than individuals who didn't, resulting, ultimately, in a situation where individuals with that trait existed rather than individuals who lack this trait. When we say, for example, that the long-necked-ness of giraffes is the result of natural selection, we are saying, roughly, that the ancestors of

---

<sup>12</sup> Street 2008: 208. See also Ruse & Wilson 1986, esp. pp. 174-9; Joyce 2006b, esp. pp. 4, 13; Joyce forthcoming, esp. p. 7; Levy 2006: 574.

giraffes were such that, those among them who had slightly longer necks than the others tended to produce more offspring, and that this led to a situation where a greater and greater proportion of the creatures that were born had long necks, which eventually led to the long-necked giraffes that exist now existing *instead of some other creatures with shorter necks*. This is a claim about the process that led to there being individuals who had the trait in question, not a claim about the process that led to these individuals having this trait.

Similarly, the EB claim is most naturally understood as a claim about the process that led to there being individuals who have the beliefs in question, not a claim about the process that led to these individuals coming to have these beliefs.

Finally, the predicative reading of EB doesn't have prima facie plausibility on its side. What would it even be for the forces of natural selection to push you, or me, or any other individual, towards certain beliefs?

My thesis in what follows is that all major EDAs exploit the predicative-quantificational ambiguity in their central claims. The predicative reading is required to generate epistemic worries about our moral beliefs, but only the quantificational reading is plausible. EDAs only sound convincing because they equivocate between these two readings. Until debunkers can provide either an argument for the predicative claim, or an argument from the quantificational claim to some epistemic problem with our moral beliefs, the realist faces no pressing objection from evolutionary considerations.

In Sections 2-3 I consider a range of different EDAs, and argue that, at least if their central reasoning is understood quantificationally, they can't generate the bad consequences for realism that they claim to.

In Section 4 I consider an argument that debunkers could use in support of the predicative reading. I argue that even if it succeeds, it won't help the debunker. The argument hinges on whether some quantificational readings can also count as predicative. I argue that if there are any such readings, they would also be epistemically inert, and so debunkers can't use this argument to avoid the problem I pose here. Even if the argument succeeds, then, it would not show that there are any readings of EB that are both plausible and have epistemic consequences for the realist.

One clarificatory point to emphasise about the scope of my argument here: my task here is not to show that there are no tenable epistemic objections to moral realism. Rather, my claim is only that debunkers have not given us any reason to think that any tenable epistemic objections stem from the evolutionary backstory of our moral beliefs.

## **Section 2: Warrant-Undermining EDAs**

In this section I consider warrant-undermining EDAs: arguments that take (1)-(3) to show that there is something epistemically wrong with our moral beliefs, that would stop them from being knowledge, even if they were true. I consider four epistemic

defects, and argue in each case that (1)-(3), understood in the quantificational sense, give us no reason to think that our moral beliefs have the defect in question.

## 2.1 First potential defect: BAD CAUSAL ORIGIN

Can the causal origins of a belief render it epistemically defective? Some debunkers argue that they can. Richard Joyce offers the following example in support of this:

Suppose there were a pill that makes you believe that Napoleon won Waterloo, and another one that makes you believe that he lost. Suppose that there were also an antidote that can be taken for either pill. Now imagine that you are proceeding through life happily believing that Napoleon lost Waterloo (as, indeed, you are), and then you discover that at some point in your past someone slipped you a ‘Napoleon lost Waterloo’ belief pill. It is not a matter of your learning of the existence of such pills and having no way of knowing whether you have ever taken one; rather, we imagine that you somehow discover beyond any shred of doubt that your belief is the product of such a pill. Should this undermine your faith in your belief that Napoleon lost Waterloo? Of course it should.’<sup>13</sup>

Matthew Bedke offers the following example:

[Suppose an agent], Bea, [...] is convinced that there is a goblin war raging all around her, where the goblins and their weapons are not composed of physical matter and not made of anything that can interact with the physical world. It really *seems* to Bea

---

<sup>13</sup> Joyce 2006: 179.

that such a goblin war rages on, she believes it on that basis, and she waxes poetic about being in touch with such a fact. Suppose brain imaging provides some causal explanation for why it is that Bea has such bizarre seemings, and so why it is that Bea believes in the goblin war. A crucial part of Bea's brain is being impinged by a tumor, and similar pressures on this region in other patients have caused similar effects. This explanation proceeds without appealing to any goblin wars. Once Bea is informed of the imaging results so she justifiably believes there is some full causal story for her seemings and beliefs, is Bea's seeming enough to support her belief that there is a goblin war as described? No.<sup>14</sup>

Debunkers like Joyce and Bedke argue that, if realism is true, our moral beliefs are in a similar position to the beliefs in these examples because their evolutionary backstory has nothing to do with the truth of these beliefs.<sup>15</sup> But are they right?

The analogy relies on the evolutionary claim being understood in the predicative way. In both the Napoleon pill case and the goblin war case, it is true of the people involved that they only believe that P because of factors that have nothing to do with P's truth. If the evolutionary claim is taken in the quantificational sense, however, the appearance of similarity with these cases disappears. On this reading, the evolutionary claim says nothing about the causal process by which individuals formed their moral beliefs - it's just a claim about the causal process that led to *there being lots of individuals* who have the beliefs in question. The fact that *this* causal process was not guided by truth doesn't look to be incompatible with the causal process that led each individual to his/her moral beliefs being an entirely epistemically respectable one.

---

<sup>14</sup> Bedke 2009: 198.

<sup>15</sup> See also Ruse 2006: 22-23; Street 2006: 13-14; Griffiths & Wilkins 2015: 106. Causal EDAs are discussed in Shafer Landau 2012: 25-32.

Now, it is often remarked that the realist owes a positive account of how it is that we are able to have any knowledge of the mind-independent moral facts they postulate. And opponents of realism may be right that it is not clear how they can meet this challenge, and that this is a significant problem. However, this is not the objection that EDAs are pushing. EDAs are saying that whatever *other* problems realists face, including perhaps the lack of a convincing story about how we grasp moral facts, they *also* face further problems that stem from the fact that our moral beliefs have an evolutionary explanation. Debunkers are claiming, then, that however it is that realists think we do come to know moral facts, evolutionary considerations show that we aren't doing it.

If I'm right, the quantificational reading of (1)-(3) doesn't succeed in showing this. Whatever kinds of causal story of belief acquisition are epistemically respectable, (1)-(3), understood quantificationally, doesn't do anything to show that individual moral believers didn't acquire their moral beliefs in one of these ways.

### **Three other potential defects: Counterfactual Profile**

Some EDAs argue that the evolutionary origin of our moral beliefs implies that they have a kind of counterfactual profile that stops them from being knowledge. One version of this complaint is that our moral beliefs don't vary across possible worlds with the moral facts, but rather with the facts about what is fitness-enhancing. Neil Levy puts the point this way:

we have no reason to think that our moral responses track genuinely moral features of the world. Instead, they are more likely to track fitness-relevant

features of our environment, which are not plausibly identified with moral features'<sup>16</sup>

There are two components to this tracking point:

**Sensitivity:** Our moral beliefs aren't sensitive to the moral facts, in the sense that we would give them up in situations where they aren't true

**Robustness:** Our moral beliefs aren't robust with respect to irrelevant facts, in the sense that we would continue to hold them in situations where they are still true - where all you vary is some irrelevant facts.

A third complaint that is sometimes made is that had facts about natural selection been different, we would have had very different, rather outlandish, moral beliefs. I discuss these three complaints separately.

One preliminary remark: these counterfactual requirements are most plausibly construed as placing restrictions only on what we do with our beliefs in *nearby* possible worlds. It would be implausible to require, in order for S to know that P, that it must be the case that *every* world in which P is false is one in which S doesn't believe P, or that *every* world in which P is true is one in which S believes it, or that S not believe anything outlandish in any possible world.

---

<sup>16</sup> Levy 2006: 574.



## 2.2 Second potential defect: LACK OF ROBUSTNESS

Sometimes debunkers appear to take (1)-(3) to show that our moral beliefs aren't held sufficiently robustly to count as knowledge. If our moral beliefs hadn't been fitness enhancing, so the argument goes, we would not have had them. By hypothesis, their fitness-enhancingness has nothing to do with their truth. So varying some feature of the beliefs that has nothing to do with their truth results in our no longer believing them. So these beliefs are not robust.

Sharon Street appears to be making this claim here:

[Since my normative judgements have] been shaped by causes such as my upbringing, cultural background, and inherited psychological tendencies; it is clear to me that *had some or all of these factors been different, I wouldn't have made the same set of normative judgments that I now make.*<sup>17</sup>

The success of this argument from robustness hinges on the following two questions:

(i) is lack of robustness an epistemic defect? And (ii) do (1)-(3) show that our moral beliefs lack robustness?

Is lack of robustness an epistemic defect? It's not entirely clear that it is. I walk into a room, and see Maria, who is wearing a red jumper and is just about to leave. And on this basis I form the belief that Maria is wearing a red jumper. If I had walked in slightly later, I wouldn't have seen her, and so I wouldn't have had this belief. But

---

<sup>17</sup> Street, Manuscript. p. 1. Emphasis mine.

nothing follows from this counterfactual about the epistemic status of my belief that Maria is wearing a red jumper.

But let's assume for argument's sake that the notion of robustness can be pinned down in a way that rules out that kind of case, and makes it plausible that lack of robustness is an epistemic defect. Do (1)-(3) show that our moral beliefs lack robustness? First, recall that if lack of robustness is a defect, it's not that our beliefs all need to be *maximally* robust - that there's *no* possible world where they're true but we don't believe them. The requirement only concerns *nearby* possible worlds. But here the possible world under consideration is not - by any stretch of the imagination - a *nearby* one.

But put this problem on hold. There is another problem: (1)-(3) doesn't even show that individual moral believers would give up their moral beliefs beliefs when non-truth-involving factors are varied. Again, the problem comes down to the predicative-quantificational ambiguity:

PREDICATIVE CLAIM: If natural selection had gone differently, such that P wasn't fitness enhancing, the individuals who in fact believe P wouldn't have believed that P.

QUANTIFICATIONAL CLAIM: If natural selection had gone differently, such that P wasn't fitness enhancing, the human species wouldn't believe that P, in the sense that different people would exist, people who had different evaluative tendencies, and didn't believe that P.

Only the predicative claim equates to a lack of robustness. If the quantificational claim is true, it doesn't bear on the robustness of anybody's belief that P. *The situation described isn't one in which we give up our moral beliefs — it's one in which we don't exist.*

Not only does the quantificational claim fail to entail the lack of robustness claim, it's hard to see how it could pose any other epistemic problems. Why should the possibility of creatures existing who don't believe that P have any bearing on whether my belief that P, or your belief that P, is knowledge? While most of us accept that whether an agent S has knowledge can depend on modal facts about what S herself might have believed in other circumstances, it is quite another thing to hold that modal facts involving *other agents* and *their* beliefs bear on whether S's beliefs are knowledge. Perhaps it is even plausible that facts about what other *actual* agents *in fact* believe might bear on whether S's beliefs are knowledge, but it is quite another thing to hold that not only actual agents, but also possible ones, are relevant.

To summarise: (1)-(3), understood quantificationally, doesn't tell us anything about the robustness of our moral beliefs. Even if lack of counterfactual robustness is a problem, and even if the world in which these beliefs weren't fitness-enhancing, counted as a nearby one, we don't have a problem here, since the situation described isn't one in which we give up our moral beliefs — it's one in which we don't exist.

### **2.3 Third potential defect: OUTLANDISH BELIEFS**

Some debunkers emphasise Charles Darwin's claim that had evolutionary history gone differently, if, say, humans had evolved more similarly to bees, '[u]nmarried

females would, like the worker-bees, think it a sacred duty to kill their brothers, and mothers would strive to kill their fertile daughters; and no one would think of interfering'.<sup>18</sup>

Sharon Street argues, similarly, that had different evaluative tendencies been fitness-enhancing (such as the tendency of lions to kill offspring that aren't their own, or the tendency of social insects to treat the interests of their group as more important than their own individual interests), 'our system of full-fledged, reflective evaluative judgments would have looked very different as well'.<sup>19</sup>

To assess the argument from outlandish beliefs, we need to consider two questions: (i) would it constitute an epistemic defect with our actual moral beliefs if we could have had outlandish moral beliefs? (ii) do (1)-(3) show that our moral beliefs have this defect?

Let's consider the first question first. It's not entirely clear what the worry is supposed to be. But here is one attempt to reconstruct the thought that it would be bad for realists if it turns out that individuals could have had outlandish moral beliefs. Realists

---

<sup>18</sup> Darwin 1871: 122.

<sup>19</sup> Street 2006: 120-121. See also Ruse & Wilson 1986: 186. Opponents of the debunkers have also largely taken this kind of consideration seriously as a potential worry: David Copp explicitly takes it to be a desideratum that any successful response to EDAs 'should support the idea that, *even if natural selection had led our moral psychology to be somewhat different from what it is*, and even we had had somewhat different moral beliefs as a result, it is likely that our moral beliefs still would have tended to approximate to the truth.' Otherwise, he says, 'there would be a sense in which it would simply be a fortunate accident that there is a tendency for our moral beliefs to approximate to the truth.' (See Copp 2008: 197-8. Justin Clarke-Doane similarly assumes that defusing the argument that 'natural selection could have led our moral psychology to be different' would involve taking this possibility seriously, and showing that even in such a case we'd still have a general tendency to approximate to the truth, since our explanatorily basic moral beliefs about the goodness of pleasure and the badness of pain would remain the same. (Clarke-Doane 2012: 320).

(at least those who think we have moral knowledge) would have to say that these outlandish moral beliefs are false. So perhaps the worry is that we could have had false beliefs, or that we could have had beliefs that are not only false, but false in a particularly ‘out-there’ way. And if that’s true of us, perhaps that casts doubt on our being reliable believers with respect to morality.

Let’s suppose for argument’s sake that if we were susceptible to outlandish moral beliefs in this way, it would render our actual moral beliefs epistemically defective. Does the reasoning in (1)-(3) show that our moral beliefs have this defect? Not if (1)-(3) is understood in the quantificational sense. What follows from the quantificational reading of (1)-(3) is only the claim that if outlandish moral beliefs had been fitness-enhancing, ***there would be creatures who had those beliefs***. It doesn’t follow that *we* - in the sense of you, and I, and everybody else - would have had those beliefs.

Again, why should the possibility of creatures existing who have very different beliefs from us pose any kind of threat to the claim that the beliefs we have qualify as knowledge? Such creatures are still *not us*. Presumably everyone will allow that it is possible for there to be creatures alive who get things wrong - who have a lot of false - even outlandishly false - beliefs with respect to a given domain. Even a domain for which realism is pretty uncontroversial.

Perhaps the argument is supposed to be an extension of a plausible view in the epistemology of disagreement, that we are rationally required to revise down our credence in P when we come across an epistemic peer who thinks that not-P. But if this is the reasoning, it rests on at least two controversial assumptions. First, it’s controversial to claim that not only are we rationally required to revise down our

credence in P when a sufficient number of our epistemic peers think that not-P, but also that we are rationally required to do this *any time it is even possible* for there to have been a sufficient number of epistemic peers who think that not-P. At the very least, such a claim would need further argument.

Second, the argument relies on the claim that these possible believers are/would be our epistemic peers. But it's not clear how this can be assumed without begging the question against the realist. Typical peer disagreement cases involve a one-off judgement, where the disagreement has few ramifications for the subjects' other beliefs. This is no accident. When the disagreement is over a single proposition that is relatively isolated from their other beliefs, there's no tension between the claim that two people disagree, and the claim that they are both relatively competent epistemic peers with respect to the subject matter of the disputed beliefs. But the more extreme the disagreement, the harder it is to maintain both that (i) both parties are relatively competent with respect to the subject matter of the disputed beliefs, and (ii) that they are epistemic peers.

If I believe that 20% of our restaurant bill would be £15.23, and you believe it would be £14.89, that's compatible with our both being pretty good at mental arithmetic. After all, even someone who is generally reliable can make a mistake.

But if I believe that murder is wrong, and that parents have a moral duty to be kind to their children, and I meet someone who thinks that mothers have a duty to kill their daughters, this doesn't look compatible with our both being pretty good at discerning moral facts. If this person turned out to be right, this wouldn't be consistent with my by and large getting things right with respect to morality at all - I would be gravely in

error (and if I am right, this person is gravely in error). We can't both be pretty good at discerning moral facts; so the assumption that this person is my epistemic peer involves denying that I am pretty good at discerning moral facts.

It's no good saying that the debunker wants to challenge the realist's claim that we have this ability. That is of course true, but that wouldn't entitle the debunker to rest her argument on the premise that we lack such an ability - rather, the point is to provide an argument that *gives us a reason* to think we lack the ability. The debunker is entitled to suspend judgment, of course, but to hold that these creatures with outlandish moral beliefs are our epistemic peers is not consistent with suspending judgement on whether we have an ability to discern moral facts. We can only take them to be our epistemic peers if we take ourselves to lack the ability to discern moral facts. So unless the debunker *already* has some reason to think that we lack the ability to discern moral facts, this argument is not going to succeed in showing that our moral beliefs are unwarranted.

To summarise, (1)-(3), understood quantificationally doesn't show that you and I, etc. are susceptible to outlandish moral beliefs. It doesn't get us the claim that you, and I, and other individuals, would have had outlandish moral beliefs if they had been fitness-enhancing. Rather, it gets us the claim that individuals would have existed who had these outlandish moral beliefs.

Moreover, it's not clear how this latter claim could bear on the epistemic status of my beliefs, or yours, or anybody else's. Even if the disagreement of an epistemic peer should make one reduce one's credence in a proposition, and even if the disagreement of *possible* epistemic peers, as well as actual ones, were relevant, we can't take these

individuals to be our epistemic peers unless we already have some reason to think that we ourselves aren't reliable believers with respect to morality. In which case the realist is already in exactly the kind of trouble that EDAs are supposed to land her in.

## 2.4 Fourth potential defect: LACK OF SENSITIVITY

Some debunkers argue that (1)-(3) shows our moral beliefs to be *counterfactually insensitive* in the sense that we'd still have these beliefs even if they weren't true.<sup>20</sup> The underlying reasoning appears to be as follows: we evolved to have these moral beliefs because they are fitness-enhancing. And so even if they weren't true, as long as they were still fitness-enhancing, we'd still have them.<sup>21</sup>

Does (1)-(3) entail that our moral beliefs are insensitive? There is much discussion of this argument in the literature, and various objections have been made to it,<sup>22</sup> but whether or not these objections are successful, the argument suffers from the same equivocation as the arguments discussed above. Read quantificationally, (1)-(3) doesn't entail that individuals only have the moral beliefs they have because they are fitness-enhancing; it entails only that people who have these moral beliefs, *whatever their reasons for believing them in fact are*, have tended to live longer and have more offspring, so that over time, more and more people exist who have them.

---

<sup>20</sup> This kind of argument is made by Joyce 2006: 183, Street 2006: 132, and Bedke 2009: 196. It is discussed in Shafer-Landau 2012, Clarke-Doane 2012, and Wielenberg 2010.

<sup>21</sup> Some might argue that the realist has a quicker defence against the sensitivity worry: if moral truths are necessary truths, then true moral beliefs will be vacuously sensitive - if there's no possible world in which they are false, there is no possible world in which we believe them falsely. However, for discussion of why this may be too quick, see Clarke-Doane 2012, and Shafer-Landau 2012.

<sup>22</sup> See Shafer Landau 2012, Clarke-Doane 2012.



So (1)-(3), in the quantificational sense, does nothing to close off the possibility that people who have these beliefs and the evolutionary advantages they bring, are such that they would give them up if they were not true. Sure, such people wouldn't have the evolutionary advantages in that counterfactual scenario, but the EB claim doesn't entail that people have traits that would be fitness-enhancing in far-flung counterfactual scenarios. It says only that the moral beliefs people *in fact* have are *in fact* fitness-enhancing.

Here's the central issue: the quantificational reading of the EB claim only tells us how the world came to contain individuals who have these beliefs. The beliefs are fitness-enhancing: having them makes you more likely to survive and reproduce. And the dispositions that underlie these beliefs are heritable, so more and more people are born who have these dispositions. But all this is entirely compatible with the picture that moral realists want to endorse - that moral facts are mind-independent features of the world that we have an ability to discern. Beyond the dispositions being heritable and the beliefs being fitness-enhancing, the EB claim, in the quantificational sense, doesn't say anything about what these dispositions that underlie our moral beliefs are. Many kinds of disposition could play this role, and (1)-(3) read quantificationally, doesn't do anything to narrow down which ones are plausible to ascribe to us. So it does nothing to rule out what the realist wants to say, which is that the disposition that plays this role is an ability to discern moral facts. Agreeing with the EB claim doesn't force realists to give up their claim that the disposition in question is a genuine ability to discern moral facts.

Again, the realist's claim might be implausible for other reasons. Claiming that we have an ability to discern mind-independent moral facts is not uncontroversial, and many objections have been raised to it - including that realists have a burden of proof in showing that this picture is to be preferred to other plausible alternatives. But this is not the objection EDAs are making. EDAs claim that in addition to whatever other problems this realist picture faces, there is a particular problem that stems from the fact that our moral beliefs have been shaped by natural selection. But if natural selection really does raise an additional epistemic problem for realism, debunkers need to show how natural selection rules out the realist's favoured picture, and it's not clear (at least from the sensitivity argument) that they can.

The three counterfactual profile arguments just can't get going if we take the EB claim in the quantificational sense. The crux of the problem is that the counterfactual profile requirements are requirements on what the individual whose beliefs you're assessing does in certain possible worlds; not requirements on what kinds of individuals certain possible worlds contain. For an individual's belief that *P* to be sensitive to *P*'s truth is a matter of there not being any nearby possible worlds in which *P* is false *and that individual believes P*. It's not a matter of there not being any nearby possible worlds in which *P* is false *and the individuals in that world believe P*. Similarly, for an individual's belief that *P* to be robust with respect to irrelevant facts, is a matter of there not being any nearby possible worlds in which *P* is true and that individual doesn't believe *P*. It's not a matter of there not being any nearby worlds in which *P* is false *and the individuals in that worlds believe P*. And finally, for an individual to not be susceptible to outlandish moral beliefs, is a matter of there not being any nearby possible worlds in which that individual has outlandish moral beliefs. It's not a matter of there not being any nearby possible worlds that contain individuals with outlandish moral beliefs.

To summarise, I've discussed four kinds of epistemic defect that debunkers have thought that our moral beliefs would have under realism, as a consequence of (1)-(3). I've argued that (1)-(3), at least when understood quantificationally, don't establish that our moral beliefs would have any of these defects. Moreover, there's reason to think that (1)-(3), read quantificationally, just can't bear on the epistemic status of our moral beliefs *in any way*. If the EB claim is read in the quantificational sense, all it tells us is that these beliefs must be the product of some kind of heritable disposition(s). But this leaves open the very portion of logical space that the realist will want to occupy: that the disposition in question is a genuine ability to discern moral facts. Unless the debunker can rule out this possibility, she will have failed to make any kind of trouble for the realist.

### **Section 3: Truth-Undermining Arguments**

The last section considered four different warrant-undermining EDAs - arguments that try to show that our moral beliefs couldn't be knowledge even if they were true. I turn now to EDAs that contest the claim that the moral beliefs people have are even true at all.

The thrust of the truth-undermining arguments is easiest to see when we consider moral believers third-personally. Then we can separate the question of whether these beliefs are true, from the question of whether the believers in question believe them.

1\*: these creatures' moral beliefs have been shaped by natural selection

2\*: natural selection is indifferent to the truth of these beliefs

3\*: these creatures' moral beliefs have been shaped by forces that are  
indifferent to their truth (from 1, 2)

One kind of way to argue from 1\*-3\* to the claim that these creatures' moral beliefs are unlikely to be true, would be to take 1\*-3\* to show that their moral beliefs are epistemically defective in one of the ways discussed in Section 2. In general, learning that someone has formed a belief in an epistemically defective way, very often gives you good reason to think that belief is unlikely to be true. As Russ Shafer-Landau puts it:

Suppose, for instance, that someone told you that there were exactly

5,422,000,000,000 fish in the world's oceans. You have no idea whether this number is even close. Then you discover that he landed on this figure by assuming that it was identical to the U.S. trade deficit for 2011. You now have all you need in order to discredit his belief. Even though you are in no position to verify the correct number.<sup>23</sup>

However, since these kinds of truth-undermining arguments rely on the ability of the EB claim to show that the beliefs it explains must be epistemically defective, they face the same problem as the warrant-undermining arguments: the quantificational reading of EB just can't generate the epistemic consequences, and the predicative reading just doesn't look plausible.

---

<sup>23</sup> Shafer-Landau (2012), p. 2.

But there is another kind of truth-undermining EDA that doesn't (or at least doesn't explicitly) proceed via the claim that our moral beliefs are epistemically defective:

Consider all the different moral beliefs these creatures could hold. Then consider that the beliefs they do hold are the ones favoured by natural selection. Add also the claim that natural selection is indifferent to whether the beliefs are true - whatever belief natural selection favours, natural selection's favouring of it has nothing to do with whether it is true. Since we know that the beliefs they have are favoured by natural selection, shouldn't that reduce our credence in the beliefs being true?<sup>24</sup>

This reasoning is faulty. This situation described is of following kind: Suppose we are interested in whether some object O has a given property P. We know that of all the different properties O could have had, P is one of many possible ones. That is, there are many many possible scenarios in which O lacks P, and of all the many possible scenarios, only a very small proportion are such that O has P. Then we learn that O has some other property, Q, which we know is entirely independent of P.

This new bit of information lends no weight to the hypothesis that it also has P. But should it *reduce* our credence in O's having P? No. There are two things to note. First, the prospects of O having P, the property we're interested in, *already look bad*, before Q is even mentioned. What makes them look bad is that all we know about the scenario we're actually in, is that it is one of many many possible scenarios, most of which are such that O lacks P.

---

<sup>24</sup> Street appears to be making this kind of argument in Street 2006: 121-2.

Second, the discovery that O has *Q* *doesn't make the prospects for O's having P any worse*. It shouldn't lower our credence that O has P. Since *Q* is entirely unrelated to P, it doesn't help, but it also doesn't make matters any worse.

Here's another situation of the same kind: Suppose someone has picked a card at random from a normal pack of cards, without showing you what it is. There are lots of different possible cards they could have picked. Suppose you are interested in whether it is an ace. Suppose now it is revealed to you that it is a heart.

A card's being a heart doesn't bear at all on whether it is an ace. It doesn't make it more likely, and it doesn't make it less likely. So the discovery that the card is a heart shouldn't *lower* your credence that it's an ace (even though it also shouldn't raise it).

Similarly, then, if the fitness-enhancingness of a moral belief doesn't bear on whether it's true, the discovery that these creatures' moral beliefs are fitness enhancing shouldn't *lower* our credence that they are also true (even though it also shouldn't raise it).

The reason why the chances of the moral beliefs being true look so low, is that the debunker is assuming that the only thing that is relevant to determining whether these creatures' beliefs are true are evolutionary considerations. And sure, if you start out with the stipulation that natural selection is the only thing we have to go on in judging whether these creatures' moral beliefs are true or not, then it is going to look random whether they'll be true. Again, however, two things are noteworthy:

First, it's this stipulation that is making the prospects of the beliefs being true look bad, and, crucially, the evolutionary considerations don't make them any worse.

Second, to assume that we're in a relevantly similar position with respect to the question of whether our own moral beliefs are true, is to beg the question against the realist. The realist thinks that with respect to the moral beliefs that humans in fact have, we are not in the position of knowing nothing that is relevant to determining whether these beliefs are true. Rather, any moral realist is going to hold that we have a way of discerning these mind-independent moral truths - not an infallible way, but a way that is good enough for us to count as often enough having moral knowledge, or at least justified or reliable moral beliefs.

Now, perhaps moral realists are not right to hold this. Again, it is often objected that realists owe - and are unable to supply - a convincing positive story about how this happens. But again, EDAs purport to do something different from that kind of objection: they claim to show that whatever view realists hold about how it is that we discern moral truths, evolutionary considerations show that we can't be doing that thing. If I'm right, this argument fails to establish this. Even if moral knowledge is deeply implausible on the realist picture - as well it might be! - the claim that our moral beliefs were selected for doesn't make it *more* implausible.

Let's take stock of the argument so far. I have been arguing that EDAs get an undeserved appearance of plausibility by equivocating between a predicative and a quantificational reading of their central complaint. (1)-(3) are plausible only if read quantificationally, but it is only by illegitimately shifting to the predicative reading that debunkers are able to trade on misleading analogies like the Napoleon pill case and

the goblin war case, or draw conclusions about what individuals would believe in counterfactual scenarios. The quantificational reading doesn't support those analogies, nor does it tell us anything about the counterfactual profile of individuals' moral beliefs. Moreover, there's reason to think that (1)-(3), read quantificationally, just can't bear on the epistemic status of our moral beliefs *in any way*. (1)-(3), understood quantificationally, fails to rule out the possibility that the disposition that underlies our moral beliefs is a genuine ability to discern moral facts.

In the next section, I consider an objection to what I have argued so far, in the form of an argument for reading the EB claim predicatively.

## Section 4 - An Objection

I argued in §1 that debunkers haven't given any reason to accept the predicative reading of the EB claim. But perhaps they can supply an argument for doing so. It is a live debate in the philosophy of biology whether natural selection can explain why individuals have certain traits, as opposed to only explaining the frequency of these traits in a population.<sup>25</sup> Karen Neander and others have argued that natural selection *can* explain the traits of individuals. If they are right, then the EB claim could be true in the predicative sense, not just in the quantificational sense.

---

<sup>25</sup> Elliot Sober argues that natural selection can only explain the frequency of traits; Karen Neander and others argue that it can also explain why individuals have certain traits. See Sober 1984 and 1995; Neander 1988 and 1995; Matthen 1999.



I argue below, however, that this would not solve the problem. Even Neander et al accept that evolutionary explanations are explanations in the quantificational sense. They just argue that they *also* qualify as explanations in the predicative sense, and that they do the latter precisely by doing the former: it is by explaining the frequency of traits that they get to play a role in explaining why individuals have these traits. But, as I shall now argue, the problems I identified in Section 2 apply to *all* quantificational readings of EB, even if they also qualify as predicative.

This becomes clear when we look at the argument that Neander et al use. The argument begins with the claim that *inheritance* can explain the traits of individuals: my green eyes are explained, in some sense, by the fact that both my parents have green eyes. The thought is that this opens the door to natural selection being able to play an explanatory role, too, for certain traits of individuals. The fact that I have opposable thumbs is explained, in some sense, by the fact that my parents have them. But a fuller explanation would involve also explaining why *they* have opposable thumbs. Such an explanation could involve pointing out that *their parents* had opposable thumbs. And of course, we could keep going back a generation, to explain the opposable thumbs of each set of ancestors. However, argues Neander:

The correct explanation does not go on in this way ad infinitum. Somewhere back along the line (where our ancestry has long ago met and pleurably mingled) there were occasional mutations or significant genetic recombinations. Here is the key point of my argument: all (or virtually all) of these alterations in the genetic code among our ancestors tended phenotypically toward opposable thumbness. (That is, they were genetic alterations which resulted in a more opposable thumb.) This is a remarkable fact in view of the randomness of these processes, and it requires explanation. Most

significantly, *it is a fact about the frequency of a genotypic and phenotypic trait in a population.*

[...] It can only be explained by direct appeal to selection for individuals with opposable thumbs.<sup>26</sup>

What's important to note is that this argument doesn't deny that the explanation is a quantificational one - it just claims that it *also counts as* a predicative one. According to Neander et al, the way in which natural selection explains the moral beliefs of individuals is *by explaining the frequency of these beliefs in the population*. The debate is over whether explaining the frequency of a trait can count as in some sense explaining (or partially explaining) why individuals have those traits. It's not a debate over whether natural selection can explain the traits of individuals in some way *other than* by explaining the frequency of these traits. All parties to this debate accept that the explanatory work natural selection can do of traits (whether in the predicative or the quantificational sense) is done, at bottom, *by explaining the frequency of traits in a population*.

But the arguments I've given in Section 2 are arguments for thinking that explanations of the frequency of our moral beliefs can't have any bearing on the epistemic status of these beliefs. And this remains the case *whether or not such explanations count as also explaining the moral beliefs of individuals*. A brief look at the different kinds of EDAs again, confirms that the arguments of Section 2 apply even if Neander et al are right.

Consider the worries about causal origin. Neander et al will agree that natural selection explains why certain moral beliefs are so prevalent among humans. But here's the crucial point: whether or not that counts as an explanation, in some sense, of why I have the moral beliefs I have, why you do, etc., it's not an explanation in the same kind of sense as the Napoleon pills or the brain tumour. So again the argument

---

<sup>26</sup> Neander 1988: 425-6.

from causal origin can't piggyback on these sorts of cases; a separate argument would be required.

Moreover, it's hard to see how the kind of backstory involved in the EB claim *could* show that our moral beliefs have an epistemically bad causal origin. Again, take whatever causal story of belief acquisition *would* be epistemically respectable; the claim that individuals acquired their moral beliefs that way - however implausible you think that would be for other reasons - isn't made any *more* implausible by the fact that what led to there being so many individuals with these moral beliefs had nothing to do with their truth.

Next consider the sensitivity worry. I argued that the EB claim does nothing to rule out the realist's claim that the dispositions underlying our moral beliefs are genuine abilities to discern moral facts, so it doesn't establish that our moral beliefs are insensitive to the moral facts. If Neander et al are right, natural selection explains, in a sense, why I have a given fitness-enhancing disposition. But a disposition that is *in fact fitness-enhancing*, doesn't have to be *fitness-enhancing in all possible worlds* - so there is no reason to think that the disposition that underlies my moral beliefs, and yours etc., is one that would produce these particular beliefs in *all* worlds where they are fitness-enhancing, including worlds in which they were false.

Finally, consider the robustness and outlandishness worries. Both are worries about what would happen in counterfactual situations where the fitness-enhancingness of beliefs is different. Vary the fitness-enhancingness of an in-fact evolved trait, and you vary which creatures reproduce, and which creatures are born. So the situation in which different moral beliefs were fitness-enhancing, is one in which different

creatures exist, not a situation in which you, I, and other individuals who in fact exist, have different moral beliefs. And the situation in which the moral beliefs we in fact have are not fitness-enhancing, is also one in which different creatures exist, not one in which you, and I lack the moral beliefs that we in fact have. Whether, as Neander claims, natural selection can in some sense explain the traits of individuals, doesn't change any of this, since Neander agrees that the way in which natural selection does this is by influencing which individuals survive and reproduce.

Now, perhaps it will be objected that my arguments here assume the essentiality of ancestry: that it is a necessary truth that each individual has the ancestors she in fact has. But my arguments don't in fact rely on this assumption. Essentiality of ancestry would be *one reason* to think that the individuals in these counterfactual scenarios are not us. So it is true that denying the essentiality of ancestry would remove one barrier to holding that they are us; but debunkers need far more than this. Debunkers need it to be the case that these individuals *definitely are us*. The problem is they haven't given any reason to think that they would be.

Perhaps, though, one might worry that if ancestry is not essential, this frees up the possibility of making the Neander-type argument in a way that makes the robustness and outlandishness worries look more persuasive. As Mohan Matthen puts it: 'Why do I have an opposable thumb? Part of the answer is: because of natural selection, there were no other genes in the pool from which to make me'.<sup>27</sup>

But then it looks like a short step from this to saying:

---

<sup>27</sup> Matthen, pp. 149-50.

(C1) if the gene pool had been composed differently, I would not have had  
opposable thumbs.

After all, if there were no opposable thumb genes in the pool, I would not have had  
opposable thumbs because there would have been no opposable thumb genes from  
which to make me.

But if C1 looks okay, wouldn't it also be okay to say:

(C2) if the gene pool had been differently constituted, I would not have the  
moral beliefs that I in fact have

and:

(C3) if the gene pool had been differently constituted, I would have different  
moral beliefs?

And C2 and C3 just amount to the robustness and outlandishness worries,  
respectively.

This argument would not be successful. The problem is that counterfactuals like C1  
don't follow straightforwardly from Matthen's claim. If there is a sense in which these  
counterfactuals are true, it's this: hold fixed that I exist, and then consider what  
properties I would have, assuming I exist, in different kinds of counterfactual scenario.  
What's doing the work in Matthen's original claim, and in C1-3 is the thought that in  
the relevant scenario, 'there [would be] no other genes in the pool from which to make

me'. But this can only get us C1-3 if we hold fixed that the scenario must be one in which I exist. A lack of opposable thumb genes in the pool doesn't guarantee that I exist without opposable thumbs - all it gets us is that *if I exist* I don't have opposable thumbs. But here's the crucial point:

(O1) if outlandish moral beliefs had been fitness-enhancing, *then if I existed* I would have outlandish moral beliefs

is a very different claim from

(O2) if outlandish moral beliefs had been fitness-enhancing, then I would have outlandish moral beliefs.

It's one thing to claim that the epistemic respectability of my moral beliefs depends on O2's being false; it's quite another to claim that it depends on O1's being false. Since Matthen's argument can only get us O1, it doesn't give us any reason to worry about the epistemic status of our moral beliefs, at least in terms of the outlandishness worry.

Similarly for the robustness worry. Matthen's argument can't get us:

(R1) if my actual moral beliefs hadn't been fitness-enhancing, I wouldn't have had them,

but only the weaker claim:

(R2) if my actual moral beliefs hadn't been fitness-enhancing, then *if I existed* I would lack these beliefs.

Only R1 amounts to lack of robustness.

So even if Matthen is right about natural selection's ability to explain individuals' traits, the way in which it does so doesn't support the outlandishness worry about our moral beliefs, nor does it support the robustness worry. And unless debunkers can supply some reason for thinking that (O2) or (R2) would entail that my moral beliefs are epistemically defective, Matthen's argument doesn't give us a reading of the EB claim that would bear on the epistemic status of our moral beliefs.

In summary, whether or not Neander, Matthen et al succeed in showing that natural selection can explain the traits of individuals, there does not appear to be any plausible reading of the EB claim that would allow (1)-(3) to generate worries that our moral beliefs are epistemically defective. My claim, either way, is that explanations of how the world came to contain individuals with these moral beliefs, are not the kind of explanations that can bear on the epistemic status of these beliefs. And this is so *whether or not* such explanations also qualify as in some sense explaining why individuals have the beliefs. If they do explain this, they don't do so in any sense that would be epistemically relevant. So even if Neander et al turned out to be right, this wouldn't affect my arguments against EDAs. But although the spirit of my claim would remain the same if Neander's argument were successful, the letter of it would require one modification. The distinction across which EDAs equivocate would not be the distinction between predicative and quantificational readings of the EB claim, but between quantificational and non-quantificational readings. The quantificational

reading would, as before, be the only reading of EB that is plausibly true, while only a non-quantificational reading would plausibly generate epistemic consequences for our moral beliefs.

There is one recent paper in the debate on EDAs that comes close to recognising the problem I have articulated here. In ‘Do Evolutionary Debunking Arguments Rest on a Mistake about Evolutionary Explanations?’, Andreas Mogensen objects to EDAs on similar grounds to those invoked here. If what I’ve argued here is correct, however, Mogensen’s arguments don’t go far enough: EDAs face a problem that is more severe than the objection Mogensen raises, in at least two respects.

First, Mogensen takes the problem with EDAs to be that they rest on controversial assumptions. He argues that EDAs ‘beg important questions in the philosophy of biology’ because they rely on Neander et al being right that natural selection can explain the traits of individuals.<sup>28</sup> He thinks, as I do, that if natural selection *can’t* explain the traits of individuals, EDAs won’t get off the ground. But Mogensen doesn’t go far enough: he allows that if Neander et al were right, EDAs would survive the objection. But if what I have argued in this section is correct, EDAs are in trouble even if Neander et al are right.

Second, Mogensen’s argument relies on assuming the essentiality of ancestry, at least for the individuals that in fact exist.<sup>29</sup> If what I’ve argued above is correct, however, EDAs have a problem even if our ancestry is not essential to us.

---

<sup>28</sup> Mogensen 2015: 1800.

<sup>29</sup> Mogensen 2015: 1804-5.



# Conclusion

My target in this paper has been arguments that take the evolutionary backstory of our moral beliefs to debunk those beliefs (either in the sense of showing them to be likely false, or epistemically defective in a way that would preclude their being knowledge even if true). I've argued that the evolutionary backstories of beliefs just can't do this kind of work.

The reason they appear to do so, is that they can sound a bit like the sorts of backstory claims that do have epistemic consequences. But once we appreciate what the evolutionary backstory of our moral beliefs amounts to (namely, an explanation of the prevalence of certain moral beliefs, by explaining why individuals who had tendencies towards them reproduced more, leading to there being more and more individuals with these tendencies), it looks highly doubtful that this is the kind of backstory that can bear on the epistemic status of these beliefs. I argued that this is so even if natural selection can explain the traits of individuals, for even if it can, the sense in which it would do so is not a sense that would be epistemically relevant.

I want to stress again that my conclusions here are entirely compatible with taking seriously the problems that moral realism faces. None of my arguments have relied on underestimating the challenge that realists face in explaining how we can have knowledge of abstract, mind-independent moral facts.<sup>30</sup> However pessimistic we

---

<sup>30</sup> David Enoch discusses this challenge in *Taking Morality Seriously*, p. 158. Enoch also thinks that EDAs are a species of this kind of challenge. I disagree, but that falls outside the scope of the present paper. If, by 'EDA's, Enoch means to refer to a kind of argument that doesn't especially have anything to do with evolution - in which evolutionary considerations don't do any work - then I haven't said anything here about EDAs in his sense. But, as I have argued, it is common to hold that evolutionary considerations *do* do work in undermining moral realism, and it is this view that has been my target here.

should be about the prospects of this challenge being met, debunkers are taking the evolutionary backstory behind our moral beliefs to be a source of additional worries. My claim has been that however bleak the prospects of moral knowledge look for the realist, the evolutionary backstory of our moral beliefs adds no further woes.

If I am right, this will undermine attempts to use evolutionary backstories to debunk beliefs in other areas too. In normative ethics, some philosophers have taken evolutionary explanations to debunk deontological intuitions.<sup>31</sup> If what I've argued here is correct, these explanations just can't do this debunking work. Equally, evolutionary explanations of normative beliefs in other normative domains, such as aesthetics, will - if what I've argued here is correct - be just as epistemically inert as the EB claim, and pose no greater threat to realism about these domains than the EB claim posed to moral realism.<sup>32</sup> Again, whatever problems these target views might face - and these problems may be serious - considerations about the evolutionary backstories of beliefs aren't among them.<sup>33</sup>

*Cambridge University, UK*

---

<sup>31</sup> Greene 2008; Singer 2005.

<sup>32</sup> Street 2006, for example, takes the evolutionary considerations she discusses to cause problems for evaluative realism generally, not just moral realism.

<sup>33</sup> I would like to thank Michael Blome-Tillmann, Paulina Sliwa, Natalja Deng, and anonymous referees at *Philosophical Quarterly* for helpful comments. For discussion of presented versions, I would like to thank audiences at New York University, and at the Moral Sciences Club and Serious Metaphysics Group, Cambridge University.

## REFERENCES

- Bedke, M. (2009) 'Intuitive Non-Naturalism Meets Cosmic Coincidence' *Pacific Philosophical Quarterly* 90 (2): 188-209.
- Clarke-Doane, J. (2012) 'Morality and Mathematics: The Evolutionary Challenge' *Ethics*, 122 (2): 313-40.
- Copp, D. (2008) 'Darwinian Skepticism About Moral Realism' *Philosophical Issues* 18: 186-206.
- Darwin, (1871) *The Descent of Man* (London: Penguin Classics, 2004).
- Enoch, D. (2010) 'The Epistemological Challenge to Metanormative Realism', *Philosophical Studies* 48(3): 413-38.
- Enoch, D. (2011) *Taking Morality Seriously* (Oxford: OUP).
- FitzPatrick, W. (2014) 'Debunking Evolutionary Debunking of Ethical Realism' *Philosophical Studies*: 1-22.
- Graber, A. (2012) 'Medusa's Gaze Reflected: A Darwinian Dilemma for Anti-Realist Theories of Value' *Ethical Theory and Moral Practice* 15: 589-601.
- Greene, J. (2008) 'The Secret Joke of Kant's Soul', in Sinnott-Armstrong (ed.) *Moral Psychology: The Neuroscience of Morality* 3: 35-79.
- Griffiths, P. and Wilkins, J. (2015) 'When do Evolutionary Explanations of Belief Debunk Belief?' in P. Sloan et al (eds.) *Darwin in the Twenty-First Century* Notre Dame, IN: University of Notre Dame Press.
- Joyce, R. (2006a) *The Evolution of Morality* (Cambridge, MA: MIT Press).
- Joyce, R. (2006b) 'Metaethics and the Empirical Sciences', *Philosophical Explorations*, 9: 133-48.
- Joyce, R. (2016) 'Evolution, Truth-Tracking, and Moral Skepticism' in R. Joyce *Essays in Moral Skepticism*, 142-58. Oxford: OUP.
- Kahane, G. 'Evolutionary Debunking Arguments', *Noûs* 45(1): 103-25.
- Levy, N. (2006) 'Cognitive Scientific Challenges to Morality', *Philosophical Psychology* 19: 567-87.
- Lillehammer, H. (2003) 'Debunking Morality: Evolutionary Naturalism and Moral Error Theory' *Biology and Philosophy* 18: 567-581;
- Matthen, M. (1999) 'Evolution, Wisconsin Style: Selection and the Explanation of Individual Traits' *British Journal for the Philosophy of Science* 50: 43-150.

- Mogensen (2015) 'Do Evolutionary Debunking Arguments Rest on a Mistake about Evolutionary Explanations?' *Philosophical Studies* 173(7): 1799-1817.
- Neander, K. (1988) 'What Does Natural Selection Explain? Correction to Sober', *Philosophy of Science* 55 (3): 422–26.
- Neander, K. (1995) 'Explaining Complex Adaptations: A Reply to Sober's 'Reply to Neander'', *British Journal for the Philosophy of Science* 46 (4): 583–87;
- Ruse, M. and Wilson, E. (1986) "Moral Philosophy as Applied Science," *Philosophy* 61: 173-92
- Ruse, M. (2006) 'Is Darwinian Metaethics Possible (And If It Is, Is It Well Taken)?' in G. Boniolo and G. De Anna (eds.) *Evolutionary Ethics and Contemporary Biology*, 13-26. (Cambridge: Cambridge University Press)
- Shafer-Landau, R. (2012) 'Evolutionary Debunking, Moral Realism, and Moral Knowledge', *Journal of Ethics and Social Philosophy* 7(1): 1-37.
- Shafer-Landau, R. (2003) *Moral Realism: A Defence* Oxford: OUP.
- Singer, P. (2005) 'Ethics and Intuitions' *The Journal of Ethics*, 9: 331-52.
- Skarsaune, K. (2011) 'Darwin and Moral Realism: Survival of the Fittest' *Philosophical Studies* 152: 229-43.
- Sober, E. (1984) *The Nature of Selection: Evolutionary Theory in Philosophical Focus* (Cambridge, MA: MIT Press).
- Sober, E. (1995) 'Natural Selection and Distributive Explanation: A Reply to Neander', *British Journal for the Philosophy of Science* 46(3): 384-97.
- Street, S. 'Objectivity and Truth: You'd Better Rethink It', Manuscript. Available here: [https://files.nyu.edu/ss194/public/sharonstreet/Writing\\_files/Paper%2012%20for%20website%20-%20Objectivity%20and%20Truth%20-%20You'd%20Better%20Rethink%20It.pdf](https://files.nyu.edu/ss194/public/sharonstreet/Writing_files/Paper%2012%20for%20website%20-%20Objectivity%20and%20Truth%20-%20You'd%20Better%20Rethink%20It.pdf)
- Street, S. (2006) 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies* 127: 109-66.
- Street, S. (2008) 'Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About', *Philosophical Issues* 18: 207-28.
- Wielenberg, E. (2010) 'On the Evolutionary Debunking of Morality', *Ethics* 120(3): 441-64.